Master in High Performance Computing

Novel Application of Neural Networks in Model Recognition for Cosmic Microwave Background Data Analysis

Supervisor(s):

Luca Heltai,
Nicoletta Krachmalnicoff
Carlo Baccigalupi

Candidate:

Farida Farsian

4th edition
2017–2018

Abstract

Over the last few years, Neural Networks (NN) indicate favorable characterizations in accuracy and performance in different scientific fields, especially Astrophysics and Cosmology. In the context of Cosmic Microwave Background (CMB) B-mode polarization observations, and in relation to the actual dominance of foreground emissions in the data, the issue of astrophysical model recognition has become severe and challenging. In this work, we propose a novel application of NNs to discern the best model describing the superposition of astrophysical and cosmological signals in the data. The latter operation represents the start of the development a novel layer of algorithms, to be exploited in a pre-processing prior to the reconstruction of the CMB B-modes. Our method is based on a fully-connected network that is trained on a set of multi-frequency CMB polarization maps and tested on different test sets in the absence and presence of noise. Considering the frequency coverage and sensitivity represented by future satellite and low-frequency ground-based probes, our NN is able to reach an accuracy above 90 % in different cases. Moreover, our method shows advantages over the widely-used method in the field, χ^2 information in terms of accuracy. Our results address the importance of including the NN-based algorithm in the foreground model recognition pipeline for the next generation of CMB observations.

Acknowledgements

I would like to express my special thanks of gratitude to my professors and collaborators, Luca Heltai, Carlo Baccigalupi, Nicoletta Krachmalnicoff for their guidance, support, deep and fruitful discussions. My sincere thanks also go to Alberto Sartori, Jacopo Rigosa, and Ivan Girotto, who I learned a lot from them during the master.

Also, I thank all of my friends and colleagues in MHPC 4th edition, who I learned many things from them, and I enjoyed their company; who without them staying in the MHPC room till (sometimes) midnight wouldn't be possible and bearable.

Moreover, I am grateful to my parents for their constant support and encouragement, as well as my partner, who I can count on his presence at any moment of need.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

This work is an interdisciplinary study with the aim of using advanced frameworks like Neural Networks (NN) in cosmology and more specifically, in the Cosmic Microwave Background (CMB) field.

The CMB radiation is the oldest electromagnetic radiation in the Universe and remnant from the very early stage of the cosmos. This radiation that fills all the space is one of the pillars of the modern standard cosmological model. The CMB and its polarization is an essential probe for understanding of Universe in the early phase. The CMB is partially linearly polarized due to Thomson scattering at the epoch of recombination [4], and its polarization state can be described by the standard Stokes parameters, Q and U, which are coordinate-dependent [5]. The CMB polarization pattern can be also decomposed into an alternative base, the B and E-modes, with odd and even behavior with respect to parity transformation, respectively [6, 7]. Unlike Q and U parameters, B and E-modes are coordinate-independent on the sphere.

Primordial Gravitational Waves (GWs) produced by the Inflationary era in early Universe are sources of the CMB B-mode anisotropies, and represent the main observational target of ongoing and future CMB probes [see 3, and references therein]. A second relevant and non-primordial source of B-mode anisotropies is represented by the gravitational lensing of CMB photons by forming large scale structure (LSS) [8]. The CMB lensing signal is fundamental for investigating the dark cosmological components of the Universe through LSS.

The GW contribution to B-modes, parametrized by its amplitude relative to primordial scalar perturbations, the tensor-to-scalar ratio, r, induces anisotropies at the degree and super-degree scale. The lensing signal dominates the B-mode spectrum at the arcminute angular scale [9].

The CMB field is known to possess a Gaussian distribution of anisotropies [10], and is characterized primarily through its angular power spectra. They have been reconstructed with great accuracy over the full sky, for the total intensity (T) and E-mode polarization, by the Wilkinson Microwave Anisotropy Probe (WMAP) [11] and Planck [12] satellites. An intense and global effort is currently ongoing towards the measurement of the B-mode polarization. Lensing B-modes have been detected for the first time by the South Pole Telescope [SPTpol, see 13, and references therein] through cross-correlation, and directly by POLARBEAR [14]. Moreover, they have been observed by the Planck satellite [15], the Background Imaging of Cosmic Extragalactic Polarization 2 (BICEP2) [16], the Atacama Cosmology Telescope (ACT) [17]. On the other hand, only upper limits exist so far for the amplitude of the cosmological GWs, corresponding to $r < 0.06$ (at 95% confidence level) [18].

In the last few years, it has become clear that one of the greatest challenges for the detection of primordial B-modes is represented by the control and removal of the diffuse emission from our own Galaxy. As a matter of fact, Galactic polarized radiation has an amplitude larger than the cosmological signal on the degree and super-degree scales, at all frequencies and in all the sky regions [see 1, 19, 20, and references therein]. In order to face this challenge and be able to extract a clean cosmological signal, future CMB probes are characterized by a multi-frequency coverage, with very high sensitivity detectors in all the frequency channels. Along this line, several observatories are currently being built. In particular, The Simons Array [SA, see 21] is being deployed, and the Simons Observatory [SO, see 22] will start operations in the early years of this decade. On the longer term, the Stage-IV network of ground-based observatories [CMB-S4, see 23], along with the Light satellite for the study of B-mode polarization and Inflation from cosmic microwave background Radiation Detection (LiteBIRD, [24]), are designed to reach an accuracy, after foreground subtraction, corresponding to the capability of detecting a B-mode signal with r as low as 10^{-3} with a high confidence level.

The set of algorithms dedicated to the removal of diffuse foregrounds from the

CMB signal is known as component separation, and consists of combining multi-frequency observations in order to reconstruct clean maps of the CMB as well as each foreground emission. In particular, typical methods for component separation are based on parametric fitting of the multi-frequency maps, where the parameters are represented by the amplitude and frequency scaling of the different foreground components [25, 26]. Therefore, a crucial aspect, which constitutes the focus of the present work, is represented by the need of an accurate modeling of the foreground emissions and how the relevant parametrization might vary across the sky, as it is clearly shown in recent and comprehensive analyses concerning proposals of future satellite missions [27]. An incorrect or inaccurate modeling of Galactic emissions could indeed lead to high residuals in the final CMB maps, preventing the measurement of the faint B-mode cosmological signal [23].

This issue can be thought as a model recognition problem, which represents one of the most important applications of Artificial Intelligence (AI). Neural Networks (NNs) and Machine Learning (ML) in general, as a subset of AI, can be very useful in Cosmology and specifically in the CMB field. In particular, NNs are non-linear mathematical tools characterized by many parameters which are able to model different problems with high complexity. For this reason, they are widely used in science. In the recent years, several works include applications in this direction, ranging from estimating cosmological parameters from dark matter [28], to real-time multimessenger astronomy for the detection of the GW signal from black hole merger [29] and weak lensing reconstruction via deep learning [30]. Recent applications, specific to CMB, include: foreground removal from CMB temperature maps using an MLP neural network [31], convolutional neural networks for cosmic string detection in CMB temperature map [32], lensing reconstruction [33] and convolutional NNs on the sphere [34].

In this thesis, we present a new NN application concerning the classification of the appropriate foreground model across the sky, identifying the physical parametrization which describes better a multi-frequency dataset in the different sky regions. This classification has to be seen as a pre-processing to the component separation phase, in order to instruct the latter with the proper functions to be exploited for the fitting. As a case study, in terms of frequency coverage, angular resolution, and sensitivity, we have considered the specifications of the complete frequency coverage of the LiteBIRD satellite [24] and the low frequencies channels of the Q and U Joint Observatory in Tenerife [QUIJOTE, see 35]. For testing our NN

model, we have focused on the analysis on the diffuse Galactic emissions which dominate the low frequency range, i.e. 70 GHz or less, in the CMB B-mode observations. Our goal is to study if a pre-processing model recognition phase is possible, and with which efficiency and accuracy.

This thesis is organized as follows: in Chapter 2 we describe how NNs work, emphasizing the aspects which are particularly relevant for this work. In Chapter 3 we introduce the diffuse Galactic foregrounds which are important for polarization, and their spectral behavior and parametric models of diffuse Galactic foregrounds. In Chapter 5 we define the NN architecture which is implemented and adopted. In Chapter 6 we study the performance and accuracy of the NN in distinguishing different foreground models distributed differently across the sky. In Section 6.4 we compare the information provided by the NN with the one from a standard approach based on the χ^2 statistics using parametric foreground removal. Finally, in Chapter 7 we bring up the discussion and conclusions.

Chapter 2

Neural Network basic concepts

The concept behind Neural Networks (NNs) is quite old even though only in the last decade converts to a very hot topic and the related literature is flourishing [36]. The reason behind this popularity is increasing the data volume and big data analysis challenge along with the advancement of technology and computational power. The idea and development of the first NNs come from biological neurons back to 50's. The first NN, so-called *Perceptron* designed by [37], by formalizing how biological neurons work.

2.1 Perceptron

The term *perceptron* was coined during the '50s and its way of working was already incredibly similar to how modern NNs work. Since the *perceptron* can be considered an ancestor of the NNs of today and they still have a lot in common, we'll describe how a simple *perceptron* works.

The whole *perceptron* can be represented by a function that takes a vector of inputs $\bar{x}$ and gives a binary output. For each input the perceptron has one internal parameter called *weight* denoted by $\bar{w}$ and a *bias* term b, that encodes the prior knowledge. The *perceptron* does one very simple thing: it *weights* the external input $\bar{x}$ with the internal parameters $\bar{w}$, it sums them up along with b and it applies to the result an *activation function* that determines the final output of the *perceptron*. Nowadays *non-linear activation functions* are the most used although there is a vast literature about them [38].

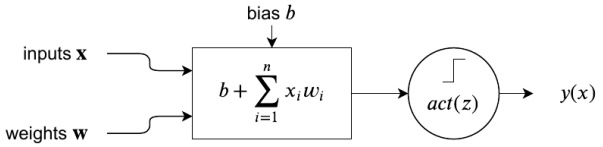

FIGURE 2.1: Schematic view of the perceptron.

The activation function of the perceptron is called Heaviside – a simple step function – and is defined as follows:

$$act(z) = \begin{cases} 1 & z \geq 0 \\ 0 & z < 0 \end{cases}$$

The output of the single perceptron will be given by

$$act(b + (\bar{x}^T \bar{w})) = act\left(b + \sum_{i=1}^{n} x_i w_i\right)$$

This definition is inspired by biological neurons and their electrophysiology: they receive some inputs, combine them and then, through the equivalent of the activation function, decide whether or not to propagate the signal [37].

2.2 Deep Neural Network (DNN)

Like in the animal brains, we can achieve interesting flexibility of the model when we link multiple perceptrons together. In particular, during the years the community started following the approach of structuring the NNs in a layered fashion [39] although, in the recent years, researchers are exploring new ways of optimizing and structuring differently the NNs to gain in performance [40], [41].

What makes a neural network "deep" is actually the number of layers between the *input layer* and the *output layer* of the network. These layers are called *hidden*. In a layered NN, the outputs of the previous layer are the inputs to the next layer. With the exception of a few more sophisticated structures, NNs usually form an acyclic graph, known as *feed-forward network*.

It can be proven that a NN is a *universal approximator* of functions [42], this means that it is possible to approximate with arbitrary precision any measurable function

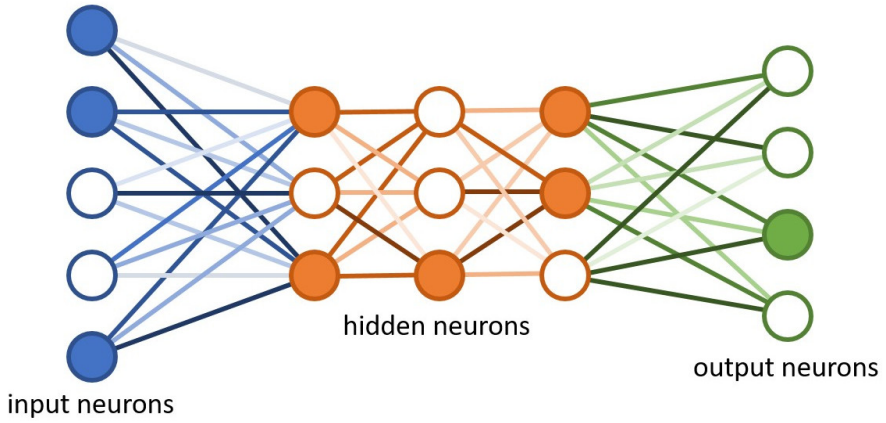

FIGURE 2.2: Schematic representation of a Deep Neural Network ans its layers.

depending on the number of neurons present in the NN. It is worth mentioning that choosing a *non-linear activation function* allows the NN to approximate even *non-linear* behaviours. This is usually a common practice, especially when using DNN.

2.3 Training of a Neural Network

Generally speaking, NNs are algorithms that recognize underlying relationships in a set of data [43]. Given a function f, that maps an input x into an output y, the goal of a NN is to find the best approximation f^* of f. In order to do that, the NN recursively applies non-linear functions to linear combinations of the input elements. In this way, the function f^* depends on several parameters θ (the coefficient of the linear combinations) which need to be optimized in order to get $f^*(\theta) \approx f$. This is done through a training set, i.e. a set of data for which the real output $y = f(x)$ is known: by computing the NN output $\tilde{y}$ for the elements of the training set, and by minimizing the distance between y and $\tilde{y}$, the best values for the NN parameters θ are found. The optimization is done numerically, usually with a *stochastic gradient descent* (SGD) method, which searches for the optimal parameters in the directions where the *gradient* is lower. The stochasticity is used to add noise to the trajectory and to avoid getting stuck into a local minima [44]. All these information are encoded in the function that determines the distance between y and $\tilde{y}$ which is called *loss function* [45].

The basic structure of NN is a *neuron*. Neurons are organized in layers; in each neuron a linear combination of all the elements of the previous layer is computed. These linear combinations are activated through a non-linear *activation function*, and the outputs of this operation become the inputs of the following layer. In the *input layer*, neurons take the value of the elements of the input x, while in the output layer the neurons take the values of the elements of $\tilde{y}$. For a general description of NN architectures can be found in [46].

The set of θ values which constitutes the best approximation of f is obtained through an iterative process, where the NN runs on the training set elements and the minimum of the loss function is found. Minimizing such *loss function* is not easy due to the high dimensionality of the problem and the underlying high non-linearity. The standard procedure that set a turning point in the world of NNs and that helps to solve efficiently this problem is called *backpropagation* [47]. The backpropagation algorithm computes, for each unit of the NN, the derivative of the error with respect to the weights in order to come up with the *gradient* of the error. Once the *gradient* has been computed, we use an *optimization algorithm* that minimizes the error and updates the weights of the NN accordingly. The values of the θ parameters are updated at each *epoch*. The number of epochs is one of the NN hyper-parameters and simply defines the number of iterations that are needed before the minimum of the loss function is reached. Given the very large number of parameters that a NN needs to optimize, over-fitting may occur; in this case, the NN approximates well the function f on the training set but it is unable to generalize to another set of data. To avoid this, a typical approach is to introduce the so-called *dropout*, i.e. a mechanism for which, in each epoch, some of the neurons of the NN are randomly switched off. This prevents the NN to rely on any specific parameter and allow it to mitigate overfitting.

Chapter 3

Polarized CMB foregrounds

The CMB radiation, in order to arrive to earth, pass through all the universe from z $\sim$ 1100 up to now, from large scale structure to our galaxy. By passing through these structure, the CMB signal will be contaminated by different radiations and sources. Any kind of this contamination is so-called CMB foreground. All the precious information that we mentioned in the last chapter are available after precise foreground cleaning. There many CMB foregrounds that one should take care of, specifically in the CMB total intensity analysis. Since the CMB B-mode polarization observation will be the main concern of the next generation of CMB experiments, in this thesis we focus on the defused CMB B-mode foregrounds which the most important ones are galactic synchrotron and thermal dust emission. Any foreground residual in the CMB B-mode polarization measurement can prohibit the detection of primordial gravitational waves, therefore, we also consider Anomalous Microwave Emission (AME) that can be polarized less than 1%.

3.1 Synchrotron emission

The synchrotron radiation is generated by cosmic-ray electrons accelerating in the Galactic magnetic field. This emission depends on the number and energy spectral index of cosmic rays electrons and the strength of the magnetic field. Thus, accurate modelling of the Galactic cosmic rays and magnetic field distribution can be useful in prediction and removal of polarized synchrotron emission. It dominates over the CMB at frequencies $\lesssim$ 70 GHz and possesses a steep Spectral Energy

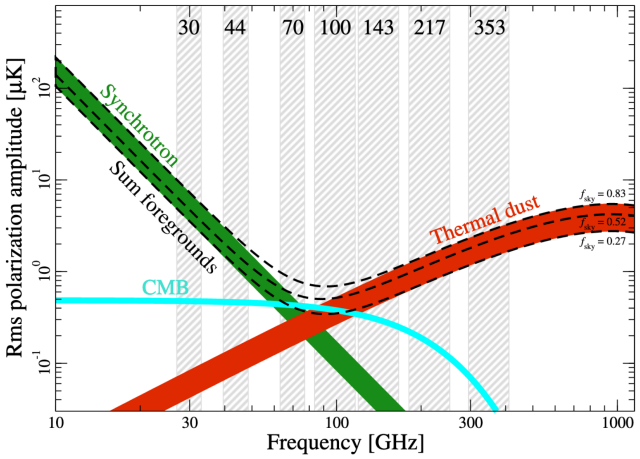

FIGURE 3.1: Polarized intensity rms amplitude of synchrotron and thermal dust emissions as a function of Planck's frequencies. The green band indicates polarized synchrotron emission, and the red band indicates polarized thermal dust emission. The cyan curve shows the CMB rms for a ΛCDM model with $\tau = 0.05$, and is strongly dominated by E-mode polarization. The dashed black lines indicate the sum of foregrounds evaluated over three different masks with $f_{sky} = 0.83$, 0.52, and 0.27. From Planck 2018 results [1]

Distribution (SED) due to the corresponding energy distribution of electrons. In Figure 3.2 Polarized intensity of synchrotron emission at 30 GHz is shown. This map describe the strong intensity of synchrotron in the Galactic plane and spur observed by Planck [1]. At first order, the synchrotron SED can be parametrized as a simple power-law in brightness temperature. Nonetheless, the energy distribution of electrons may be responsible for a curvature in the SED, which departs from a pure power-law. The general model for synchrotron emission can be written as:

$$T(\hat{n}, \nu) = A_s(\hat{n}) \left(\frac{\nu}{\nu_0}\right)^{\beta_s(\hat{n}) + C(\hat{n}) \log(\nu/\nu_0)} , \tag{3.1}$$

where A_s is synchrotron amplitude at the pivot frequency ν_0, β_s is the synchrotron spectral index, and C parametrizes SED curvature. In general, all quantities are functions of the sky direction $\hat{n}$. The synchrotron spectral index has a typical value $\beta_s \approx -3$, with a variation between -2.98 and -3.12 in the sky, on the degree scale [48]. In another recent work, the synchrotron spectral index variation has been found to be in the range between -2.5 and -4.4, with a mean value of $\beta_s \simeq -3.2$; this has been obtained by considering low frequency channels from 2.3 to 33 GHz, combining radio observations by the S-band Polarization All Sky Survey (S-PASS, see [49]), WMAP and Planck data [20].

Non-zero curvature is suggested by cosmic ray energy spectrum at frequencies

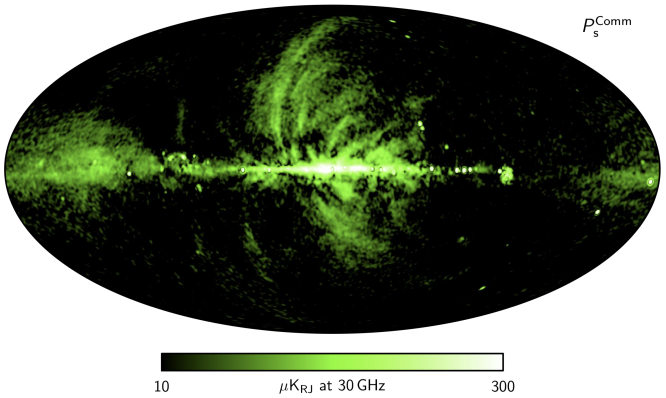

FIGURE 3.2: Planck Commander 2018 polarized synchrotron amplitude map at 40' FWHM resolution

above 23 GHz in total intensity, resulting in $C = -0.052 \pm 0.005$ [50]. [20] have derived an upper limit to the curvature value in polarization: the reported value is between 0.07 and 0.14 depending on the considered sky region and angular scales.

3.2 Thermal dust emission

Polarized thermal dust emission [see 51, and references therein] comes from interstellar dust grains which are mostly made of graphites, silicates, and PAHs (Polycyclic Aromatic Hydrocarbons) and they tend to align perpendicularly to the Galactic magnetic field, therefore emitting partially linearly polarized radiation. Dust grains are heated by starlight, and possess a modified black body SED, known as the grey body, with a temperature T_d with values around 20 K and varying across the sky. The SED is also described by a multiplicative emissivity correction ν^{β_d}, which determines the deviation from a pure black body, with β_d assuming values around 1.6 and a variation between 1.53 and 1.67 across the sky. Thermal dust emission dominates the polarized sky radiation at frequencies $\gtrsim 70$ GHz [see 1, and references therein]. In the Figure 3.3 the distribution of polarized intensity of thermal dust emission can be seen across the sky; this map was observed by Planck [1]. The analytic form of the brightness emission of the SED can be written as:

$$T(\hat{n}, \nu) = A_d(\hat{n}) \left(\frac{\nu}{\nu_0}\right)^{\beta_d(\hat{n})} B(\nu, T_d(\hat{n}))\,, \tag{3.2}$$

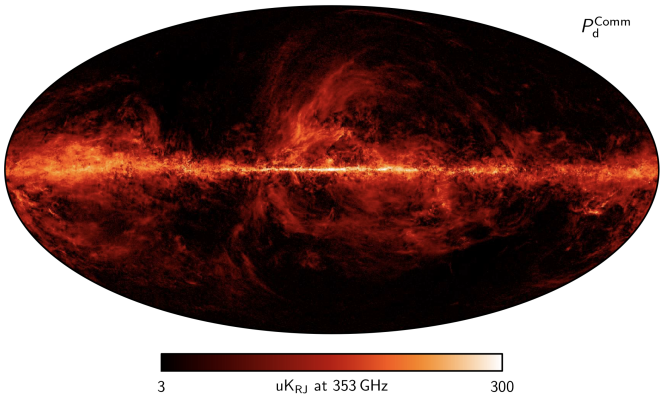

FIGURE 3.3: Planck Commander 2018 polarized thermal dust amplitude map at 5' FWHM resolution

where A_d defines the dust amplitude varying across the sky at the pivot frequency ν_0, and B represents the standard black body spectrum at the temperature T_d and frequency ν [51].

The aforementioned values for the dust spectral index β_d and T are based on this simplistic assumption that thermal dust emission is generated by one population of dust grains. There are different modellings of dust grain and populations which can fit the spectra. One of the famous models are consist of two populations of dust; therefore the model has two spectral index $\beta_{d\,1,2} = (1.67, 2.70)$ and two dust temperature $T_{1,2} = (9.4, 16)$ [52].

3.3 Anomalous Microwave Emission (AME)

In total intensity, the AME has been observed by the QUIJOTE telescope and the Planck experiment in the frequency range $\approx$ 10-60 GHz [53]. A possible explanation of this emission is represented by the spinning of the dust grains, which rotate at GHz frequencies and emit electric dipole radiation if they have an electric dipole moment [54], or magnetized dust grains and free-floating ferromagnetic material [55]. The AME SED is expected to exhibit a bell shape form, characterized by a peak at around 30 GHz. If AME is polarized, its polarization fraction must be very small, at the level of a per cent [53]. The QUIJOTE, in [56], for example, has constrained the AME polarization to be $< 2.8\%$ with 95% confidence level in the Perseus molecular complex. In another paper [57] only QUIJOTE put the upper limit of AME polarization $< 0.39\%$ and by combining the data with WMAP, it tightened the constrain to $< 0.22\%$ for the W43 molecular complexes. Note that

the aforementioned limits are measured for specific regions and cannot be applied to the whole sky. Remazeilles et al. [58] have shown neglecting 1% polarized AME can bias the extracted r value, particularly for satellite missions.

The parameterization of the considered model for AME SED is based on Ali-Haimoud et al. [59] paper. The spinning dust grains with angular velocity ω of electric dipole moment μ can radiate as follow:

$$P = \frac{2}{3}\frac{\mu_{\perp}^2 \omega^4}{c^3}\,, \tag{3.3}$$

where P is the radiation power and $\mu_{\perp}$ is the perpendicular component of μ to ω. This power is emitted at the frequency $\nu = \omega/2\pi$. The emissivity of electric dipole radiation per Hydrogen (H) atom can be calculated through:

$$\frac{I_\nu}{n_H} = \frac{1}{d\pi}\int_{a_{min}}^{a_{max}} da \frac{1}{n_H}\frac{dn_{gr}}{da} 4\pi\omega^2 f_a(\omega) 2\pi \frac{2}{3}\frac{\mu_{a\perp}^2 \omega^4}{c^3} \tag{3.4}$$

where $\omega = 2\pi\nu$. The term $\frac{1}{n_H}\frac{dn_{gr}}{da}$ determines the grain size distribution function which gives the number of dust grains per unit size per H atom, $\mu(a)$ is the electric dipole moments as a function of grain size and $f_a(\omega)$ is the angular velocity distribution function which depends upon the grain radius and environmental condition. This function is calculated for a cold neutral medium in the simulations we adopt.

In this work, we adopt the standard model of the AME, constituting of simulated polarized maps with thermal dust polarization angles and nominal AME intensity. We deployed the implemented model in `Python Sky Model (PySM)` [1] publicly available package which generates the full-sky simulation in intensity and polarization [60]. The AME model in `PySM` makes use of `SpDust2` code [59, 61] to calculate the nominal AME model and dust polarization angles (γ_{353}), are calculated from the Planck `Commander` 2015 thermal dust Q and U maps at 353 GHz [62]. The assumption of complete mixture of small and big grains leads to consider the same angles as thermal dust. The AME polarization can be written as:

$$Q_{ame} = f I_\nu cos(2\gamma_{353}), \qquad U_{ame} = f I_\nu sin(2\gamma_{353}). \tag{3.5}$$

[1] https://github.com/bthorne93/PySM_public

Where f is plorization fraction. In this work, we have considered a global 2% polarization fraction, within the limits observed by WMAP, Planck, and QUIJOTE in Perseus.

Chapter 4

Simulations

In this chapter we describe the set up adopted to simulate the sky maps, related to Galactic foreground emissions, used to train and test our NN. As anticipated, we focus on low frequency foregrounds, and we consider all the frequency channels covered by the future LiteBIRD satellite [24] plus the two lowest frequency bands with specifications of the QUIJOTE telescope [35]. Our results are conservative in this sense, because more powerful low frequency observations are being planned [63] and would results in more low frequency channels to be combined with LiteBIRD, and with more sensitivity. On the other hand, in this work we choose to see which results are achievable with the existing data. The corresponding frequencies, together with sensitivities and angular resolutions for all the considered channels are summarized in Table 4.1.

Therefore, the sky emissions included in our simulations are CMB, Galactic synchrotron, thermal dust and polarized AME. All the components are simulated using the `PySM`.

In particular, the CMB maps are generated as random Gaussian realizations of the Planck best fit Λ Cold Dark Matter (ΛCDM) power spectrum [12]. For dust, we have used the PySM template, rescaled as a modified blackbody, as in Equation (3.2), with constant spectral index $(\beta_d, T_d) = (1.54, 20K)$. For synchrotron, we have considered two different models. In the first one, the template is extrapolated in frequency with a simple power-law model. The spectral index is spatially varying, considering a Gaussian distribution with mean value β_s = -3 and standard deviation equal to 0.2. In the second case, a curvature is included in the synchrotron SED, with a constant value of $C = -0.052$, as indicated by Kogut

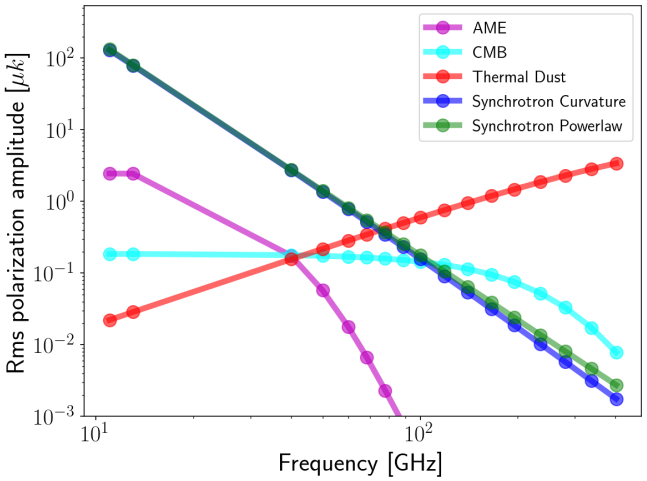

FIGURE 4.1: Polarized intensity rms amplitude as a function of frequency and different foreground component models which is used in our simulation. The plot refers to a sky fraction $f_{sky} = 78\%$ with 4 degrees pixel gridding, and units are in brightness temperature.

[50] with 23 GHz as the pivot frequency; this setup is also compatible with the recent analysis by Krachmalnicoff et al. [20]. Finally, as explained in the following Sections, we have also included, in some specific cases, the AME polarized signal, assumed to have a 2% polarization fraction. The noise is simulated uniformly in the sky, through Gaussian realizations with standard deviations given by the parameters listed in Table 4.1. In Table 4.2 we have shown the summary of considered foreground models and their parameterizations. As an illustration of the relative relevance of the various components, in Figure 4.1 we plot the rms of their polarized intensity, in brightness temperature units and gridding the sky with 4 degree pixels, for all the sky emissions and frequencies considered in this work. We have applied Planck 2018 component separation common mask in polarization with $f_{sky} = 78\%$.

Experiment	Frequency [GHz]	Sensitivity [μK-arcmin]	FWHM [arcmin]
Quijote	11.0	840.0	55.2
	13.0	840.0	55.2
LiteBIRD	40.0	36.1	69.2
	50.0	19.6	56.9
	60.0	20.2	49.0
	68.0	11.3	40.8
	78.0	10.3	36.1
	89.0	8.4	32.3
	100.0	7.0	27.7
	119.0	5.8	23.7
	140.0	4.7	20.7
	166.0	7.0	24.2
	195.0	5.8	21.7
	235.0	8.0	19.6
	280.0	9.1	13.2
	337.0	11.4	11.2
	402.0	19.6	9.7

TABLE 4.1: Frequencies and instrumental specifications for QUIJOTE and LiteBIRD. The values are consistent with recent studies, [2], [3], respectively.

Foreground models	Parameterization
Synchrotron power-law	$\mu(\beta_s) = -3$, $\sigma(\beta_s) = 0.2$
Synchrotron curvature	$\mu(\beta_s) = -3$, $\sigma(\beta_s) = 0.2$, C = -0.052
Thermal dust	$\beta_d = 1.54$, $T_d = 20K$
AME	$f_p = 2\%$

TABLE 4.2: Summary of the foreground models considered in this work. The parameterization is based on Equation 3.1 for synchrotron, Equation 3.2 for thermal dust and Equation 3.5 for AME. $\mu(\beta_s)$ and $\sigma(\beta_s)$ are the synchrotron spectral index mean and standard deviation, respectively. f_p represents the polarization fraction.

Chapter 5

Neural Network architecture

In this work, we have used NNs to recognize the actual parametrization of Galactic foregrounds in the sky. There exist several NN architectures. In this work we make use of the so-called *fully connected* ones. The architecture in this network is such that all the neurons, in one layer are connected to the neurons in the next layer. We have built the NNs in the *Keras*[1] environment, which is a *Python* library, with *Tensorflow*[2] backend. We have considered two NN architectures, which correspond to the problems we want to analyze, as described in the following.

5.1 Architecture for Binary classification

In our problem, the input of the NN are vectors of dimension 2×17. Each element of this vector represents the amplitude of the sky signal in a given pixel at the different considered frequencies (17 in total) for one of the polarization Stokes parameters. The two vectors of 17 elements each for Q and U are then stacked together to get the 34 elements long input vector.

As the purpose of this work is to solve a classification problem (assigning each pixel in the sky to a specific foreground model), the output of the NN is a vector where each element gives the probability that the input pixel belongs to any of the considered classes (models). The dimension of the output vector depends on how many possible sky models are considered, as explained in the following Sections.

[1] https://keras.io
[2] https://www.tensorflow.org

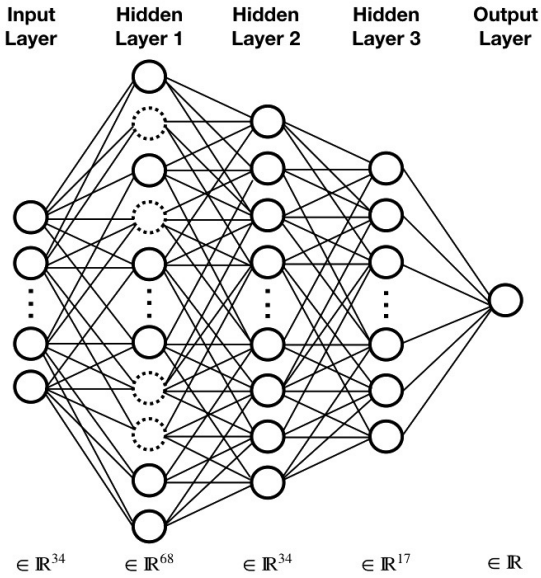

FIGURE 5.1: Schematic NN architecture used for binary classification: each circle represents a neuron, and the dashed circles indicate the application of dropout to a layer.

In the first considered case, we have trained the NN to perform a binary classification, meaning that its goal is to assign to each pixel in the sky one out of two possible foreground models. As we specified above, the NN input layer has dimension of 34, after that 3 hidden layers are present, including 68, 34, and 17 neurons each, with *tanh* as an activation function. In order to prevent overfitting, a dropout layer with a dropout rate = 0.5 is applied on the layer with the largest number neurons. Since we are in the case of binary classification, the output layer, activated with a *sigmoid* function, has, in this case, dimensionality 1, corresponding to the probability of the input to belong to the first class. Figure 5.1 shows the schematic architecture of our binary classifier. The loss function is defined as a binary-crossentropy function: $L = -(zlog(p) + (1 - z)log(1 - p))$, where p is the predicted probability for each input to belong to the specific class and z is the binary indicator associated to the two classes (0 or 1). We have used *Adadelta* optimizer with learning rate = 1.0 which is implemented in Keras. Adadelta is an adaptive learning rate method which dynamically changes the learning rate based on a moving window of gradient updates. Therefore this method needs no manual tuning of the learning rate during the learning and turns out to be robust to noisy gradient information.

5.2 Architecture for Multi classification

As we explain in the following Sections, we have also considered a case where the NN has to distinguish among four different sky models. Due to the enhanced complexity with respect to the binary classification, we increase the number of layers and neurons accordingly. In this case, the NN has 5 hidden layers with 272, 136, 68, 34, and 17 neurons, with *tanh* activation function. As before, a dropout layer with a dropout rate = 0.5 is applied to the first hidden layer with 272 neurons. The output layer is a multi-classification, with *softmax* function as activation. A Sparse-Categorical-Cross-entropy is chosen as loss function, corresponding to $L = -\sum_{c=1}^{M} z_{o,c} log(p_{o,c})$, where M is the number of categories for classification, p is the predicted probability for specific observation (o) of category c, and z represents the correct class indicator for that observation (o). The same optimizer as the binary classification is considered.

5.3 Hyper-parameters

The values of the hyper-parameters describing the architecture of a NN apparatus is usually determined empirically. That is the case of the number of layers and the number of neurons per layer. A large number of these quantities ensure performance, at the expense of computational efficiency and speed. Usually, large values of hyperparameters are chosen and progressively reduced while keeping the performance stable, reaching minimum value which is then frozen in the NN apparatus. Moreover, the model is prone to overfitting problem by having large but not necessary number of neurons and layers. In our work, we have tried several NN configurations, and have selected, for both the cases of binary or multi-classification, the architecture which showed the best performance with the least number of parameters to be optimized during training. See [46] and references therein for a general description of the hyper-parameter definition and derivation for NNs.

Chapter 6

Results

We now discuss the results of model recognition for low frequency foregrounds via NNs, both in binary and multi-model classification. The analysis is entirely based on simulated polarization maps, where the signal information is given via the Q and U Stokes parameters. We study noisy and noiseless simulated maps at LiteBIRD and QUIJOTE frequencies, as anticipated. In the following Sections we describe the results for the different test cases we consider.

As typical for NNs we distinguish between a training and test set. The training set prepares the NN, and the test case corresponds to simulated data. When noise is considered on the test set, we take two different approaches: first when the training is unchanged, and second when we include the noisy data in the training set and we discuss the improvements allowed by training on noisy sets. We intentionally alter the training set with respect to the `PySM` used in the test set, by multiplying the signal in each pixel by a random Gaussian distributed modulation factor with 30% standard deviation. In this way, the training set is substantially different, pixel by pixel, from the test. The training sets are prepared at high resolution, corresponding to $N_{side} = 1024$, while for the test set, $N_{side} = 16$. For all simulations, full sky is considered and no mask is applied. The summary of the different case which we tried and its NN accuracy is written in Table 6.1.

6.1 Foreground model recognition via binary classifications

We first use the binary classifier described in Section 5.1 to distinguish between two different foreground models. In particular, in the first case we train the NN in order to understand whether low frequency data are fitted better by a synchrotron model which does or does not include curvature of the spectral index (see Equation 3.1). Next, we focus on the case in which the synchrotron emission is described by a pure power-law and the NN is trained to recognize the presence of polarized AME.

6.1.1 Synchrotron curvature

We have trained the NN with four sets of simulated multi-frequency maps. Each set of maps consists in 34 skies, i.e. 17 frequencies for Stokes Q and U emissions. In each set we have included the emission coming from the CMB, polarized thermal dust and synchrotron simulated as described in Chapter 4. In two sets of maps, the synchrotron emission is scaled in frequency with a pure power-law, while in the remaining two a curvature is added to the spectral index. We have considered a different random realization of the CMB emission for each set of maps, as well as a different realization of the synchrotron spectral index spatial variation, which is taken from a Gaussian distribution with mean -3 and standard deviation 0.2. The synchrotron curvature in the two sets of maps is constant, with a value of $C = -0.052$, and 23 GHz as the pivot frequency. All the maps have been simulated at $N_{side} = 1024$, meaning that in total we have about 5×10^7 vectors, each of which consists of 34 elements, which are used for training. Among these, we have randomly selected 20%, which are not used for optimizing the NN weights, but as a validation set, as it is typically done for validating the performance of a NN. The size of the training set has been chosen in order to find the optimal balance between NN performances and computational costs. Since we have considered all the pixels in the sky maps for training the NN, and given the high level of non-stationarity of the Galactic signals, the vectors in the training set cover a very large dynamic range, of about four orders of magnitude. As it is done in preparing the data for NN training, we normalized each input vectors in the range between -1 and 1 as follows: the minimum and maximum value for each input vector are

computed; the minimum value is subtracted to the vector elements, and the result is divided by the difference between maximum and minimum. In order to further generalize the training set and make it substantially different from the test one, we have shifted the amplitude of each Galactic component. In particular, we have applied a multiplication to both the templates of synchrotron and thermal dust (at 23 and 353 GHz respectively): in each template, each pixel in Q and U is multiplied by a random value drawn from a Gaussian distribution with standard deviation equal to 30% of the amplitude of the pixel itself. The multi-frequency maps are then obtained by applying the correct frequency scaling to these modified templates.

In Figure 6.1 we show the training history with the accuracy reached by the NN as a function of epochs. Since we are working on a classification problem, in this case the accuracy represents the percentage of elements in the training (or validation) set which are classified correctly. We recall that the NN outputs the probability for each input pixel to belong to each considered class and that each pixel is assigned to the class that has the highest probability.

Once the NN is trained, we can apply it to the test set. In particular, we have built test maps, by making use of the PySM library, that include CMB, synchrotron, and thermal dust. Maps of the test set have been generated at $N_{side} = 16$ and without the modulation of the foreground templates in order to make our test set considerably different from the training one. In some regions, the synchrotron emission has been scaled in frequency with a simple power-law, in others, we have modified the SED by including a running parameter of the spectral index. An example of a test set map is reported in Figure 6.2: in the pixels belonging to the red regions the synchrotron SED is a pure power-law, while in the blue region a curvature is added. The color scales in Figure 6.2 report the output of the NN, i.e. the probability that each pixel belongs to the correct class. In particular, pixels shown with darker colors are those where the NN assigned the correct class, while pixels with lighter colors are those where the NN has missed the right foreground model. For sake of clarity, in the right panel of Figure 6.2, we show, in white, the pixels where the NN has made an incorrect prediction. The achieved accuracy (i.e. the percentage of correctly classified pixels) is about 98%. We have tried different combinations of patterns for synchrotron power-law and curvature in the sky, assessing that the accuracy reached by the NN is stable and does not depend on the considered sky configuration.

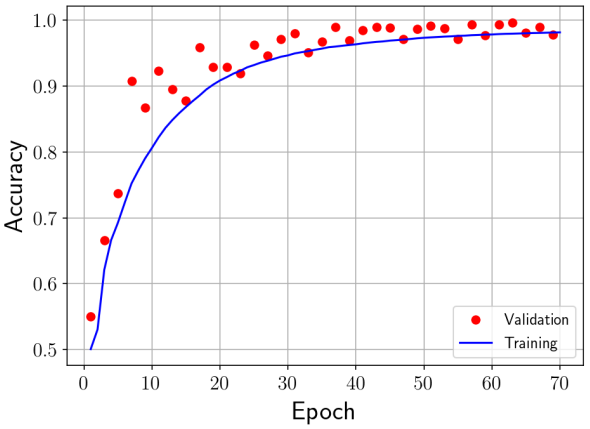

FIGURE 6.1: NN accuracy across training with respect to epochs for binary classification between synhrotron with and without curvature, in the noiseless case.

We have also investigated the physical properties of those pixels where the NN assigned the wrong model. In particular, we have found that when the relative amplitude of the synchrotron emission over dust is small, the NN has the tendency to misclassify the model. This happens for example in the region near Galactic coordinate (230°, +40°) where the synchrotron amplitude is known to be extremely weak, or on the Galactic plane where dust emission is very bright. We have quantified this effect in Figure 6.3, where we show the fraction of misclassified pixels as a function of the relative amplitude of synchrotron over dust emission. In particular, we have considered a map at $N_{side} = 256$ (corresponding to about 7.8×10^5 pixels) where we have scaled the synchrotron emission with a pure power-law on the whole sky. For each pixel, we have computed the synchrotron over dust amplitude at the frequency of 11 GHz and for the total polarized intensity. We have applied a binning on this ratio such that in each bin we have the same number of pixels (about 1600). A threshold corresponds to each bin, and we have counted the ratio of misclassified pixels over the total number of pixels with $\log(A_{synch}/A_{dust})$ below the threshold. The results in Figure 6.3 show that when the synchrotron over dust amplitude is small, the fraction of misclassified pixels increases, up to about 38%, while for the pixels where synchrotron emission is high compare to dust, the faction of misclassified pixels decreases dramatically.

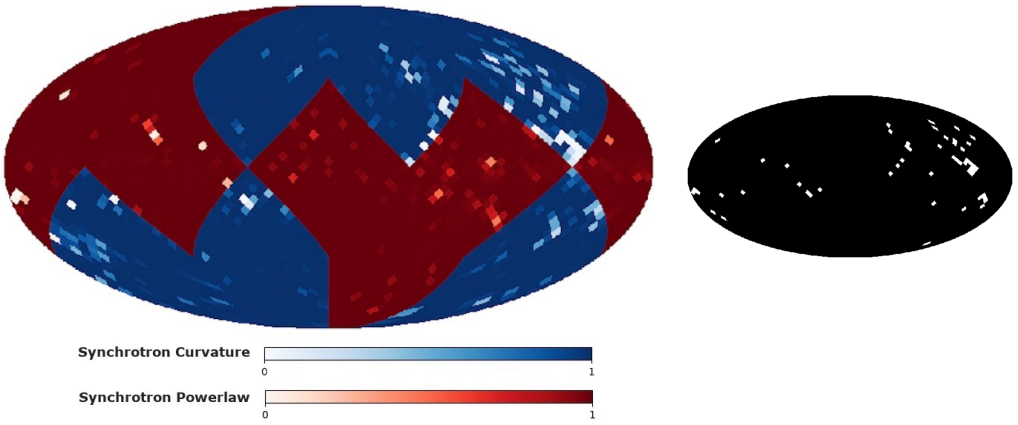

FIGURE 6.2: Left panel: NN prediction on a test set map for binary classification of the Galactic synchrotron with (blue regions) and without curvature (red regions) in the ideal case of noiseless maps. The color bar shows, for each pixel, the probability to belong to the correct class, as assigned by the NN. Lighter pixels are those where the incorrect model has been assigned. Right panel: for sake of clarity, correct (black) and incorrect (white) pixels are also shown with a binary color scale.

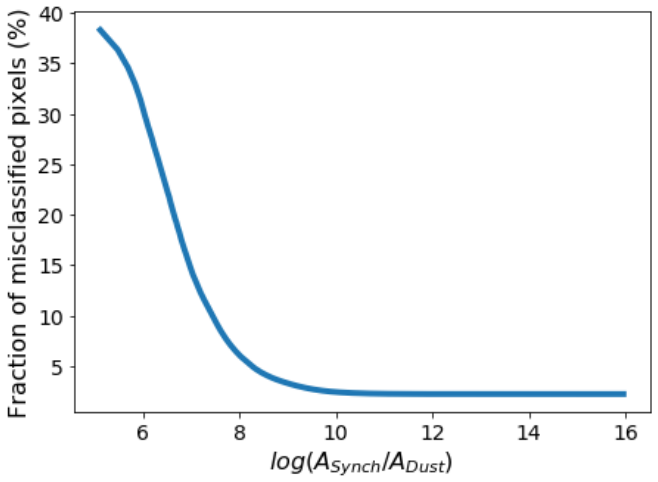

FIGURE 6.3: Faction of misclassified pixels as a function of the relative amplitude of synchrotron over dust at 11 GHz for the case of binary classification of the Galactic synchrotron with or without curvature.

6.1.2 Synchrotron and AME

We have used the same NN architecture developed for binary classification with the goal of identifying those pixels where AME polarized radiation is present in the sky. The two models considered in this case correspond therefore to Galactic synchrotron with a pure power-law SED, or synchrotron plus polarized AME component with the specifications described in Chapter 4.

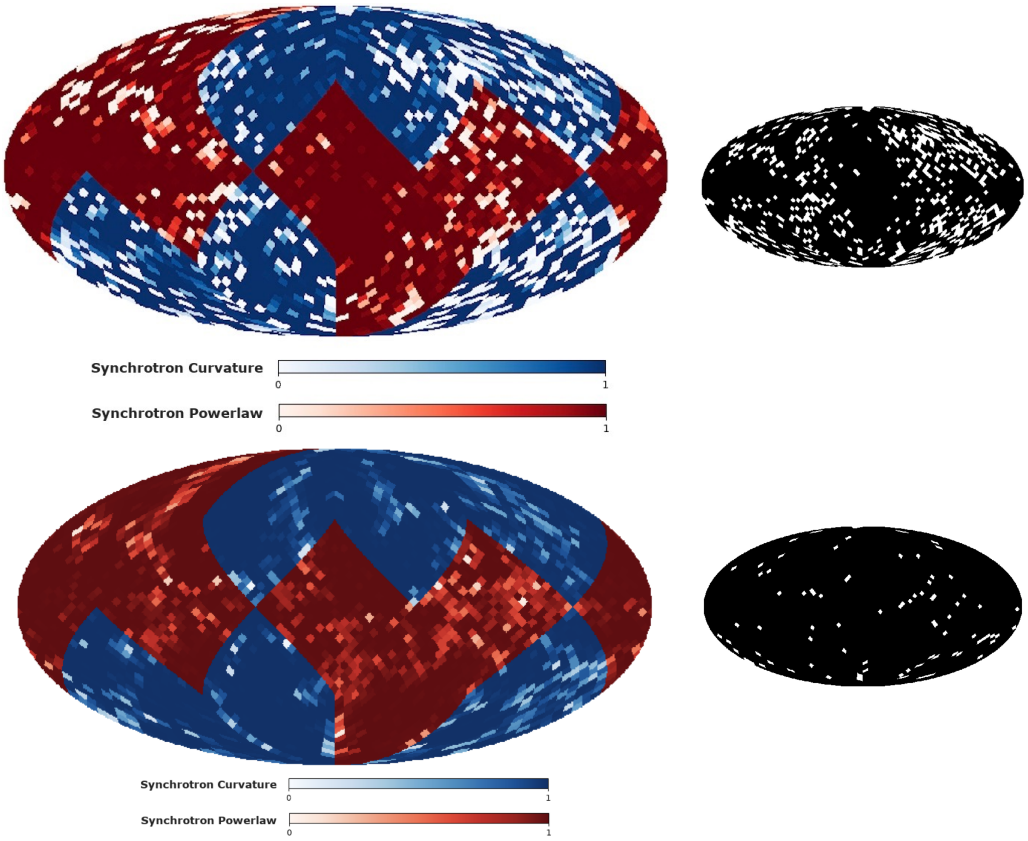

FIGURE 6.4: The effect of including the noise in the training set. The color scales for the considered models are the same as Figure 6.2. The upper panels indicate the NN accuracy on noisy test set and lower panels show the NN accuracy on the same noisy test set after re-training with 100 noise realizations at $N_{side} = 16$; As before, white pixels in the right panels are those where the incorrect model is indicated by the NN with the higher probability.

For what concerns the training, we have followed a procedure analogue to the one presented in the previous Section. The training consists of four sets of maps; in two of them we have simulated the sky emission by considering the presence of CMB, synchrotron and thermal dust radiation, while in the remaining two we have also included polarized AME. As before, the total number of vectors in the training set is about 5×10^7 and the templates of foreground emissions (dust, synchrotron and AME) have been modified by applying the multiplication factor as described in the previous Section. Results are presented in Figure 6.6, where AME is present are shown in green. In the ideal noiseless case, the NN is able to correctly classify the foreground model in about 97% of the cases. We highlight that pixels where the NN fails in classifying correctly the foreground models are those where the AME emission is faint with respect to the synchrotron one. In Figure 6.7 we report the fraction of misclassified pixels as a function of the relative

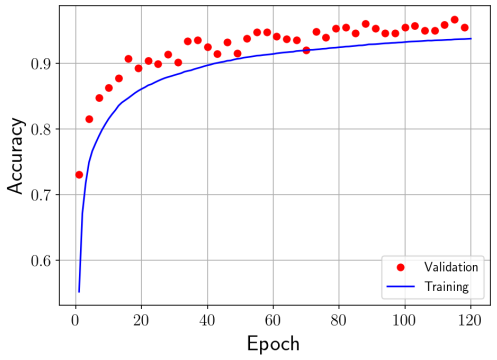

FIGURE 6.5: NN accuracy across training with respect to epochs for binary classification between synhrotron power-law and AME, in the noiseless case.

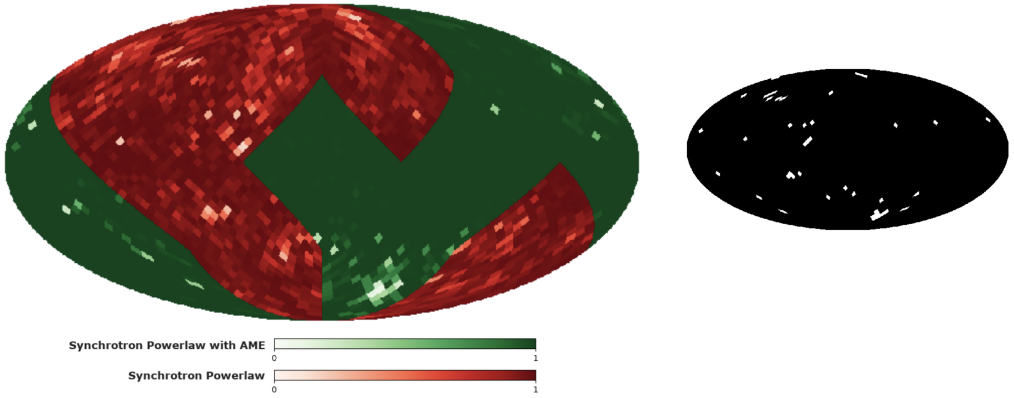

FIGURE 6.6: Left panel: NN prediction on test set for the binary classification for power-law synchrotron only (red regions) and without AME (green regions). The color bar shows the NN probability assigned to the corresponding and correct synchrotron model across the sky. Right panel: White pixels are those there the incorrect model is indicated by the NN with the highest accuracy. These results are for the noiseless case.

amplitude of AME over synchrotron at 40 GHz (the frequency closest to the AME peak), similarly to what we have done for Figure 6.3. The results show that, as expected, the smaller AME amplitude is compared to synchrotron, the higher is the fraction of misclassified pixels, up to about 40%.

6.2 Multi-model classification

We now extend the study performed so far and consider a more complex case in which the NN is trained to classify four different foreground models in the simulated sky. In this case we have used the NN architecture described in Section 5.2. As before, we have built our simulated maps by including CMB and thermal dust,

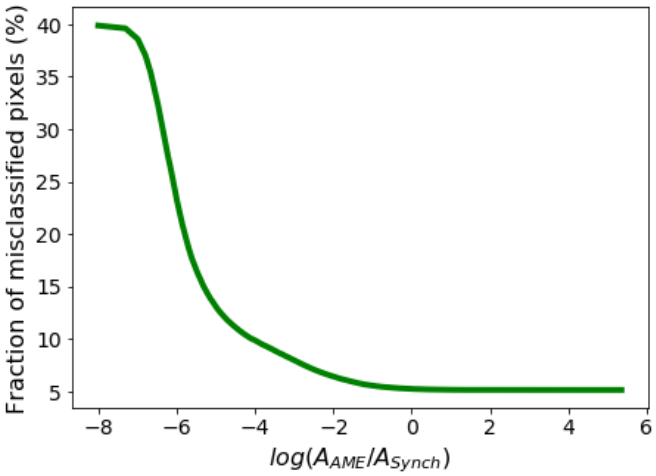

FIGURE 6.7: Fraction of misclassified pixels as a function of the relative amplitude of AME over synchrotron at 40 GHz, for the case of binary classification of the Galactic synchrotron with or without AME.

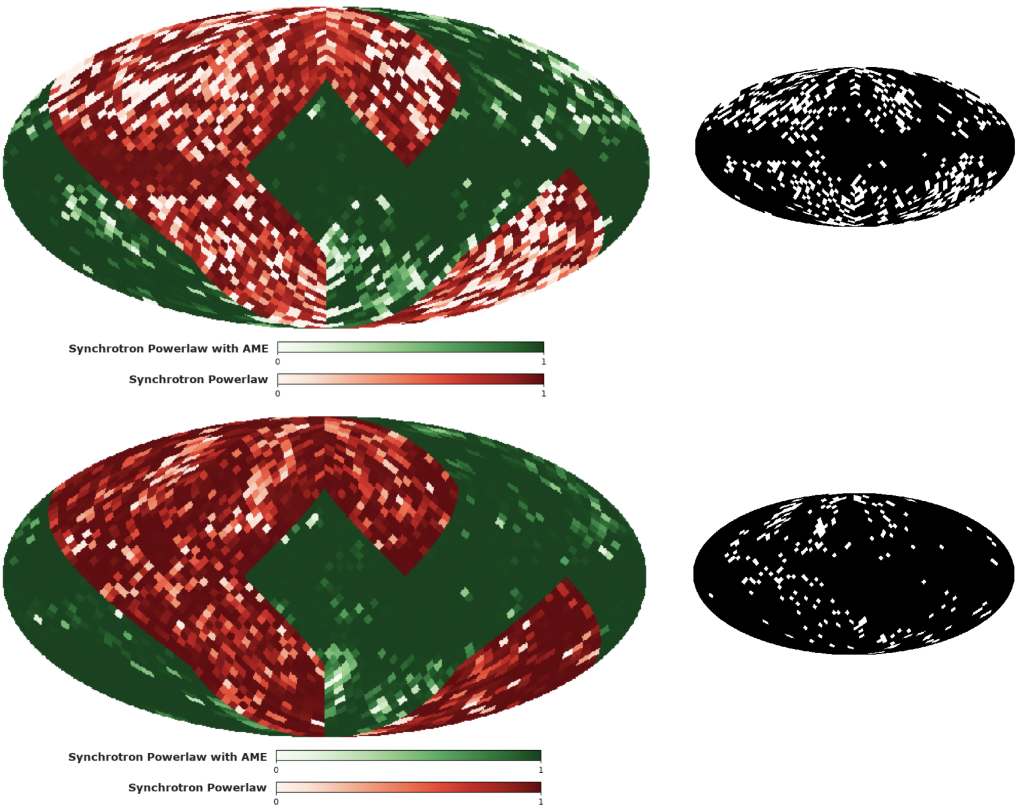

FIGURE 6.8: The effect of including the noise in the training set. The color scales for the considered models are the same as Figure 6.6. The upper panels indicate the NN accuracy on noisy test set and lower panels show the NN accuracy on the same noisy test set after re-training with 100 noise realizations at $N_{side} = 16$; As before, white pixels in the right panels are those where the incorrect model is indicated by the NN with the higher probability.

Sky models	Accuracy on training set	Accuracy on test set
Pure power-law & Curvature	99%	98%
AME & Pure power-law	93%	97%
AME & Pure power-law & Curvature	87%	93%

TABLE 6.1: Accuracy on training and test sets of the NN for different sky models in the basic configuration without noise.

while the low frequency foregrounds include synchrotron with or without a curved SED and possibly AME.

The training set has been generated from four sets of maps as before, for a total of about 5×10^7 vectors used for optimizing the NN weights. The training history is shown in Figure 6.9: the NN reaches about 87% of accuracy on the training set after 220 epochs. It is worth noticing that as a result of the enhanced complexity in the simulations, the NN training takes more time to optimize weights, reaching convergence in about 220 epochs.

Results on a test map are shown in Figure 6.10. In this case the sky is divided into four different regions, corresponding to the four models that the NN has to classify: synchrotron with a pure power-law SED (red), synchrotron with running of the spectral index (blue) and presence of polarized AME (green when AME is added to the synchrotron power-law model and purple when it is added to synchrotron with curvature). As before, color bars report the probability obtained by the NN that a given pixel belongs to the correct class, with lighter colors showing pixels where the NN has been assigned with the incorrect foreground model. The reached accuracy on the test set is at the level of about 93% and as before it does not depend on the specific pattern of models in the sky.

In Table 6.1 we report a summary of the performance of the NN in the different considered configurations. We notice that in some cases the accuracy reached on the test set is higher than the one on the training set, as it could happen as a consequence of having exploited dropout during training.

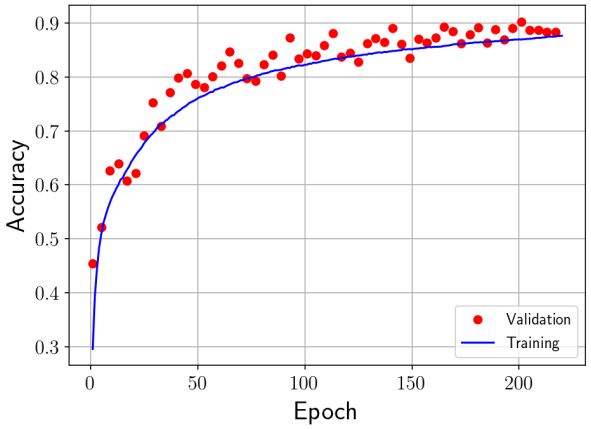

FIGURE 6.9: NN accuracy as a function of the training epochs for multi-class classification between synchrotron with or without curvature and with or without AME, in the noiseless case.

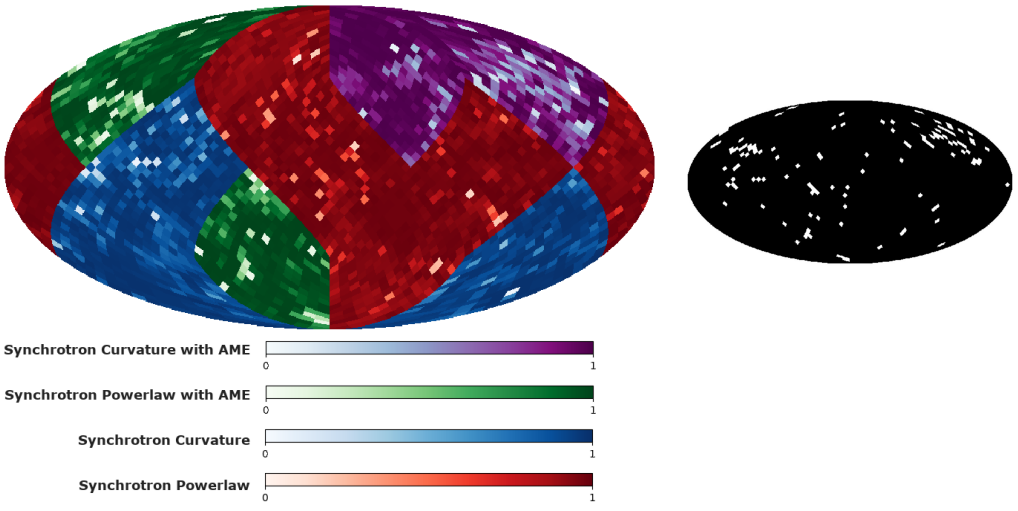

FIGURE 6.10: Left panel: NN prediction on a test set map for the multi-class classification for pure power-law synchrotron with (green regions) and without AME (red regions), or for synchrotron with curvature with (purple) and without (blue) AME. The color bars show the NN probability assigned to the correct model. Right panel: white pixels are those where an incorrect model is indicated by the NN. These results are for the noiseless case.

6.3 Classification in presence of noise

We have tested the performances of our NNs when instrumental noise is present on maps. In particular, we have considered the specification of the LiteBIRD and QUIJOTE experiments, with the sensitivities reported in Table 4.1 and uniform white noise distribution across the sky.

Our first approach has been to change only the test sets, by adding noise on the test maps, but keeping the weights of the NNs unchanged, therefore with the values optimized with the noiseless training. The first column of Table 6.2 reports the accuracy reached on the test sets for the three classification schemes we considered: binary classification for synchrotron models, presence of AME, and multi-classification. For the binary classification, we reached acceptable accuracy; While the accuracy drops significantly, reaching about 68% in the more complex multi-classification case.

In order to get better results, we have trained the NN with noise in the training set. We have considered two different approaches. In the first one, we have added one noise realization on the multi-frequency maps used previously as the training set. We have then taken the NN trained previously on noiseless data, and performed a second phase of training with the noisy training set. In this way, the NN shows a remarkable improvement in accuracy, being able to reach $\sim 90\%$ on the test set for the multi-classification. In the second approach, we have built new training sets, consisting in 100 maps for each model at low resolution ($N_{side} = 16$), resulting in 400 sets of maps included in the training set, corresponding to more than 1 million pixels. Similarly to the previous case, the accuracy is pretty high, at the level of about 93%, proving that, during training, the NN is able to learn the noise properties and take those into account during the model classification.

In Figure 6.11 we show the results on the noisy test map for the multi-model classification, for the case in which the training has been done with noiseless simulations (upper panels) and the one where the training set was obtained from low resolution maps (lower panels). A summary of all the results is reported in Table 6.2.

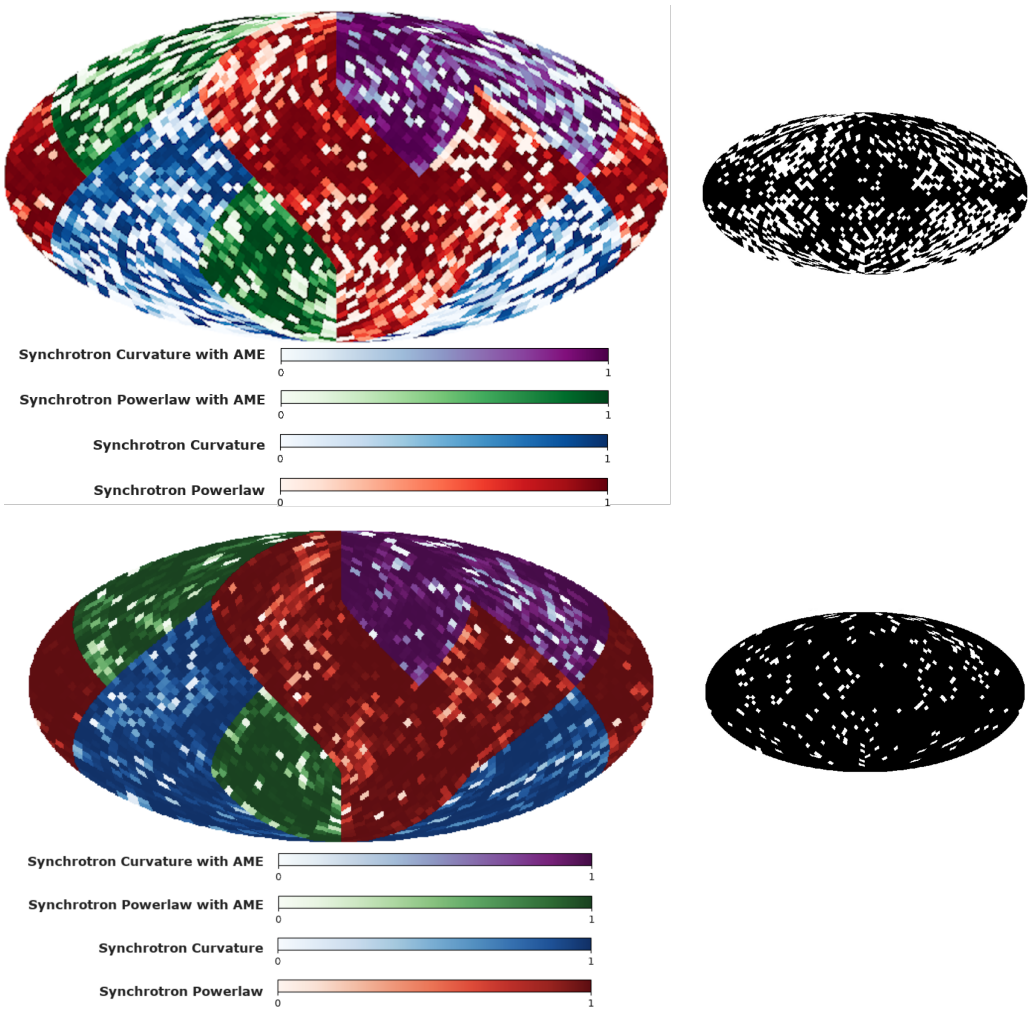

FIGURE 6.11: Effect of including the presence of noise in the training set. The color scales for the considered models are the same as Figure 6.10. The upper panels indicate the NN accuracy when noise is added only to the test set, lower panels show the NN accuracy on the same noisy test set after re-training the NN with 100 noise realizations at $N_{side} = 16$; As before, white pixels in the right panels are those where the incorrect model is indicated by the NN.

6.4 Comparison with the χ^2nformation

In this Section, we compare quantitatively the information retrieved via our NN apparatus with the ordinary goodness of fit represented by a χ^2 test following a parametric component separation analysis. Here we exploit the approach developed by Stompor et al. [25] which is currently used for quantifying the science outcome of future B-mode probes [3]. We refer to these papers for further details on this approach, limiting ourselves to the definition of quantities of relevance for

Sky models	Test set accuracy for noiseless data training	Re-training accuracy with $N_{side} = 1024$, 1 noise realization	Re-training accuracy with $N_{side} = 16$, 100 noise realizations
Pure power-law & Curvature	82%	95%	97%
AME & Pure power-law	78%	92%	94%
AME & Pure power-law & Curvature	68%	90%	93%

TABLE 6.2: Accuracy of the NN for the binary and multi-model classification in the presence of noise with different approaches for training.

the present work. The data model is usually written as

$$d_p(\nu) = \sum_c a_p^c(\nu) s_p^c + n_p(\nu) \equiv \mathbf{A}_p s_p + n_p \,, \tag{6.1}$$

where d_p contains measured signal at each frequency ν and sky direction p, summed over all components whose amplitude is written as s_p^c; $\mathbf{A}_p$ is the mixing matrix which contains the parametric SED model to fit, depending in principle on the sky direction, and n_p represents the noise. The component separation process consists in obtaining an estimate $\tilde{s}_p = \mathbf{W}_p d_p$ of the components, by means of a kernel operator $\mathbf{W}_p$, given by

$$\mathbf{W}_p \equiv (\mathbf{A}_p^T \mathbf{N}_p^{-1} \mathbf{A}_p)^{-1} \mathbf{A}_p^T \mathbf{N}_p^{-1}, \qquad \mathbf{N}_p \equiv {n_p}^T n_p \,, \tag{6.2}$$

where $\mathbf{N}_p$ represents the noise correlation matrix; the kernel operator is the result of the maximization of the likelihood

$$-2 \log \mathcal{L} = -\sum_p (d_p - \mathbf{A}_p s_p)^T \mathbf{N}_p^{-1} (d_p - \mathbf{A}_p s_p) \,, \tag{6.3}$$

which is valid in the case in which the noise is block diagonal, i.e. correlations are allowed between Stokes parameters in a given pixel only.

The χ^2 is defined as

$$\chi^2(p) = \sum_{\nu=1}^{N_{band}} \left(\frac{d_\nu - s_\nu(p)}{\sigma_\nu(p)} \right)^2 , \tag{6.4}$$

Where $\sigma_\nu(p)$ represents the uncertainty due to the presence of noise. The corresponding approach to component separation has been implemented into the publicly available code called ForeGround Buster (`FGBuster`)[1], which we adopt in the rest of the work for calculating the χ^2 after component separation, using the same input maps used so far for the NN. We restrict this analysis to the classification in the simplest cases of pure power-law or curved SED for synchrotron, i.e. the first case analyzed in the previous Section, in the binary classification mode. We run `FGBuster` on the skies used to test the NN in the presence of noise, and calculate the χ^2 accordingly. For all the pixels we fit two different models: in one case, the parameters to fit with `FGBuster` are synchrotron, dust amplitudes and synchrotron spectral index, while in the other case, in addition to those, we also fit for synchrotron curvature. Since the parameterization of two synchrotron models is different, in order to have a fair comparison between the two χ^2 tests, we have computed the reduced χ^2 taking into account the degrees of freedom.

From the reduced χ^2 we compute the probability for each pixel to belong to the correct model that we show in the upper panel of Figure 6.12. As usual, darker colors indicate the pixels where thanks to the χ^2 computation we retrieve the correct model, while lighter colors are for those pixels where the classification is wrong. We compare the results obtained from the χ^2 with those of the NN (lower panel of 6.12, in the case where we have re-trained the NN with 100 noise realization at $N_{side} = 16$ (see Section 6.2). The reached accuracy calculated from reduced χ^2 is at the level about 73%, while the NN is able to distinguish two models with of 97% accuracy. This clearly shows the gain in using a NN for model recognition.

In Figure 6.13 we also show the difference between the χ^2 values computed in each pixel for the two different cases (with or without fitting for curvature) across the sky. As it is clear, the difference between the two reduced χ^2 is very close to zero in the region where the sky signal is low (greenish regions at intermediate and high Galactic latitudes). These are the regions where the χ^2 analysis leads to a higher probability of misclassification of the foreground model, due to the low

[1] https://github.com/fgbuster/fgbuster

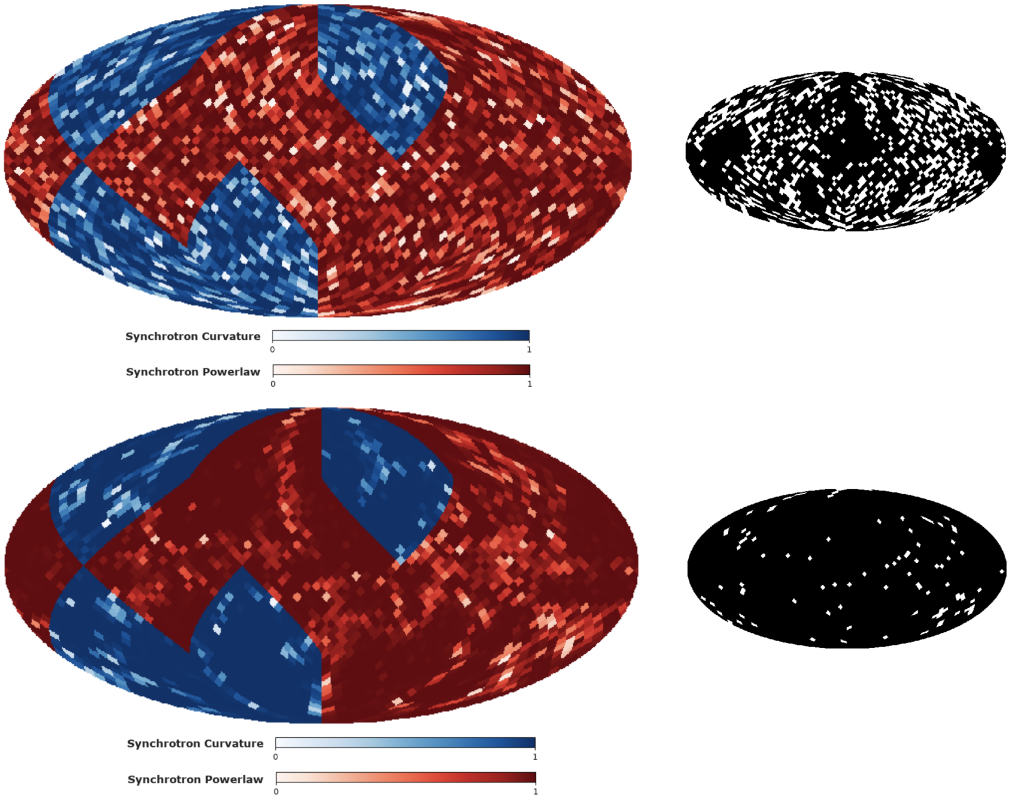

FIGURE 6.12: Comparison of χ^2 analysis with NN prediction in the presence of noise on test maps. Upper panels: on the left, we report the probability for each pixel to belong to the correct model as obtained via the χ^2 approach. In the red regions, the correct model is represented by a synchrotron power-law SED, while in the blue region a curvature is present. Lighter pixels are those where the χ^2 analysis leads to a wrong model classification (also shown in white in the right panel). Lower panels: same as the upper panels but in this case the probability has been obtained via the NN approach. This comparison shows the advantage of using a NN approach, leading to a correct classification on about 97% of the pixels with respect to about 73% when the χ^2 information is used.

signal-to-noise ratio. The same effect does not seem to affect the NN classification so strongly.

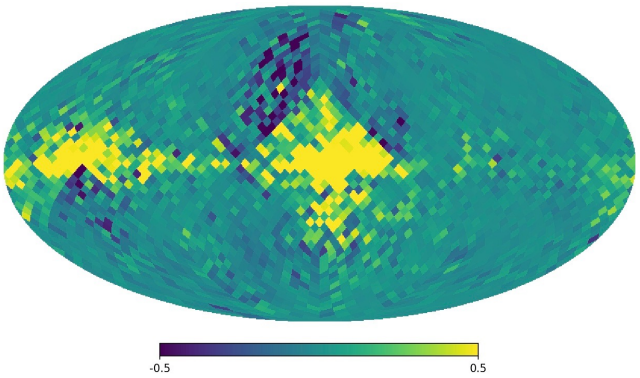

FIGURE 6.13: Reduced χ^2 difference for each pixel, obtained when the fit is done considering pure power-law SED for synchrotron and when the curvature is included. Pixels at intermediate and high Galactic latitudes (in green) are those where χ^2 is unable to distinguish between the two models.

Chapter 7

Conclusion

In this work, we start to investigate the relevance of NN in recognizing the physical properties of the diffuse linearly polarized emission from our own Galaxy at microwave frequencies, which represents the main astrophysical contaminant to the measurement of the CMB B-mode polarization sourced by GWs in the early Universe. The problem is particularly challenging and urgent, due to scientific relevance of the cosmological signal, and the difficulty in disentangle it from the much brighter foreground emission.

Foreground cleaning is usually performed via parametric fitting, which implies the necessity of identifying the physical parameters describing the foreground model in each portion of the sky, fitting and marginalizing them on the basis of a suitable multi-frequency coverage. On the other hand, foreground physical properties and model do vary in the sky, in a manner which is currently only partially revealed by observations, and yet crucial, because the right parametrization of them is necessary to perform a good fitting and to prevent the presence of large foreground residual in the CMB maps which could bias the scientific results.

In the present work, we study the possibility to identify the right physical parametrization of foregrounds, varying across the sky, in a pre-foreground cleaning phase. We do it with NNs, trained on simulations, and applied to test cases. We focus on the properties of Galactic synchrotron and AME, which have a rich phenomenology, resulting in possible different parametrization across the sky. We take care of making the simulations substantially different from observations, by explicitly and microscopically altering the training set with respect to the test one, at each resolution element. We find a good performance of the NN in recognizing the

right parametrization of foregrounds, which achieve better results with respect to a standard χ^2 test on the goodness of fit, making our results interesting and suitable for future studies.

The combination of the simulations based on the specification of the QUIJOTE telescope and the LiteBIRD satellite, with a good coverage of the relevant frequencies, are analyzed in the binary and multi-class classifications modes, i.e. when two and four models have to be recognized in the sky, respectively. In all cases, the rate of success in recognizing the right foreground model is equal or larger than 90%. This is true even in the case where four foreground models have to be recognized, namely pure power-law SED with or without curvature for synchrotron, with and without AME. We compare the NN information concerning model recognition with the χ^2 distribution following a parametric component separation assuming a given model, implemented and run through the publicly available `FGBuster` code. We find that the NN perform better wit respect to the χ^2, in particular at intermediate and high Galactic latitudes.

We believe that these results are quite interesting, and a promising first step into the construction of a model recognition layer of data analysis in B-mode CMB measurements. Further lines of investigation concern the extension to other foreground models as well as the inclusion of possible realistic systematic effects.

Bibliography

[1] Y. Akrami et al. Planck 2018 results. IV. Diffuse component separation. 2018.

[2] Josquin Errard, Stephen M. Feeney, Hiranya V. Peiris, and Andrew H. Jaffe. Robust forecasts on fundamental physics from the foreground-obscured, gravitationally-lensed CMB polarization. *JCAP*, 1603(03):052, 2016. doi: 10.1088/1475-7516/2016/03/052.

[3] Paolo Campeti, Davide Poletti, and Carlo Baccigalupi. Principal component analysis of the primordial tensor power spectrum. *JCAP*, 1909(09):055, 2019. doi: 10.1088/1475-7516/2019/09/055.

[4] Wayne Hu and Martin J. White. A CMB polarization primer. *New Astron.*, 2:323, 1997. doi: 10.1016/S1384-1076(97)00022-5.

[5] Uros Seljak and Matias Zaldarriaga. Signature of gravity waves in polarization of the microwave background. *Phys. Rev. Lett.*, 78:2054–2057, 1997. doi: 10.1103/PhysRevLett.78.2054.

[6] Marc Kamionkowski, Arthur Kosowsky, and Albert Stebbins. Statistics of cosmic microwave background polarization. *Phys. Rev.*, D55:7368–7388, 1997. doi: 10.1103/PhysRevD.55.7368.

[7] Matias Zaldarriaga. Nature of the E B decomposition of CMB polarization. *Phys. Rev. D*, 64:103001, 2001. doi: 10.1103/PhysRevD.64.103001.

[8] Antony Lewis and Anthony Challinor. Weak gravitational lensing of the CMB. *Phys. Rept.*, 429:1–65, 2006. doi: 10.1016/j.physrep.2006.03.002.

[9] N. Aghanim et al. Planck 2018 results. VIII. Gravitational lensing. 7 2018.

[10] Y. Akrami et al. Planck 2018 results. IX. Constraints on primordial non-Gaussianity. 5 2019.

[11] C. L. Bennett, D. Larson, J. L. Weiland, N. Jarosik, G. Hinshaw, N. Odegard, K. M. Smith, R. S. Hill, B. Gold, M. Halpern, and et al. Nine-yearwilkinson microwave anisotropy probe(wmap) observations: Final maps and results. *The Astrophysical Journal Supplement Series*, 208(2):20, September 2013. ISSN 1538-4365. doi: 10.1088/0067-0049/208/2/20. URL http://dx.doi.org/10.1088/0067-0049/208/2/20.

[12] R. Adam et al. Planck 2015 results. I. Overview of products and scientific results. *Astron. Astrophys.*, 594:A1, 2016. doi: 10.1051/0004-6361/201527101.

[13] D. Hanson et al. Detection of B-mode Polarization in the Cosmic Microwave Background with Data from the South Pole Telescope. *Phys. Rev. Lett.*, 111 (14):141301, 2013. doi: 10.1103/PhysRevLett.111.141301.

[14] P. A. R. Ade et al. A Measurement of the Cosmic Microwave Background B-Mode Polarization Power Spectrum at Sub-Degree Scales from 2 years of POLARBEAR Data. *Astrophys. J.*, 848(2):121, 2017. doi: 10.3847/1538-4357/aa8e9f.

[15] P. A. R. Ade et al. Planck intermediate results. XLI. A map of lensing-induced B-modes. *Astron. Astrophys.*, 596:A102, 2016. doi: 10.1051/0004-6361/201527932.

[16] P. A. R. Ade et al. Measurements of Degree-Scale B-mode Polarization with the BICEP/Keck Experiments at South Pole. In *53rd Rencontres de Moriond on Cosmology La Thuile, Italy, March 17-24, 2018*, 2018.

[17] Thibaut Louis et al. The Atacama Cosmology Telescope: Two-Season ACTPol Spectra and Parameters. *JCAP*, 1706(06):031, 2017. doi: 10.1088/1475-7516/2017/06/031.

[18] P. A. R. Ade et al. BICEP2 / Keck Array x: Constraints on Primordial Gravitational Waves using Planck, WMAP, and New BICEP2/Keck Observations through the 2015 Season. *Phys. Rev. Lett.*, 121:221301, 2018. doi: 10.1103/PhysRevLett.121.221301.

[19] N. Krachmalnicoff, C. Baccigalupi, J. Aumont, M. Bersanelli, and A. Mennella. Characterization of foreground emission on degree angular scales for CMB B-mode observations - Thermal dust and synchrotron signal from Planck and WMAP data. *Astron. Astrophys.*, 588:A65, 2016. doi: 10.1051/0004-6361/201527678.

[20] N. Krachmalnicoff et al. S-PASS view of polarized Galactic synchrotron at 2.3 GHz as a contaminant to CMB observations. *Astron. Astrophys.*, 618: A166, 2018. doi: 10.1051/0004-6361/201832768.

[21] A. Suzuki et al. The POLARBEAR-2 and the Simons Array Experiment. *J. Low. Temp. Phys.*, 184(3-4):805–810, 2016. doi: 10.1007/s10909-015-1425-4.

[22] Peter Ade et al. The Simons Observatory: Science goals and forecasts. *JCAP*, 1902:056, 2019. doi: 10.1088/1475-7516/2019/02/056.

[23] Kevork N. Abazajian et al. CMB-S4 Science Book, First Edition. 2016.

[24] A. Suzuki, P. A. R. Ade, Y. Akiba, D. Alonso, K. Arnold, J. Aumont, C. Baccigalupi, D. Barron, S. Basak, S. Beckman, and et al. The litebird satellite mission: Sub-kelvin instrument. *Journal of Low Temperature Physics*, 193(5-6):1048–1056, May 2018. ISSN 1573-7357. doi: 10.1007/s10909-018-1947-7. URL http://dx.doi.org/10.1007/s10909-018-1947-7.

[25] R. Stompor, Samuel M. Leach, F. Stivoli, and C. Baccigalupi. Maximum Likelihood algorithm for parametric component separation in CMB experiments. *Mon. Not. Roy. Astron. Soc.*, 392:216, 2009. doi: 10.1111/j.1365-2966.2008.14023.x.

[26] Ben Thorne, Jo Dunkley, David Alonso, Maximilian H. Abitbol, Josquin Errard, J. Colin Hill, Brian Keating, Grant Teply, and Edward J. Wollack. Removal of Galactic foregrounds for the Simons Observatory primordial gravitational wave search. 5 2019.

[27] M. Remazeilles et al. Exploring cosmic origins with CORE: B-mode component separation. *JCAP*, 04:023, 2018. doi: 10.1088/1475-7516/2018/04/023.

[28] Siamak Ravanbakhsh, Junier Oliva, Sebastien Fromenteau, Layne C. Price, Shirley Ho, Jeff Schneider, and Barnabas Poczos. Estimating Cosmological Parameters from the Dark Matter Distribution. 2017.

[29] Daniel George and E. A. Huerta. Deep Neural Networks to Enable Real-time Multimessenger Astrophysics. *Phys. Rev.*, D97(4):044039, 2018. doi: 10.1103/PhysRevD.97.044039.

[30] Arushi Gupta, José Manuel Zorrilla Matilla, Daniel Hsu, and Zoltán Haiman. Non-Gaussian information from weak lensing data via deep learning. *Phys. Rev.*, D97(10):103515, 2018. doi: 10.1103/PhysRevD.97.103515.

[31] H. U. Norgaard-Nielsen and H. E. Jorgensen. Foreground removal from CMB temperature maps using an MLP neural network. *Astrophys. Space Sci.*, 318: 195–206, 2008. doi: 10.1007/s10509-008-9912-6.

[32] Razvan Ciuca, Oscar Hernandez, and Michael Wolman. A convolutional neural network for cosmic string detection in cmb temperature maps. *Monthly Notices of the Royal Astronomical Society*, 08 2017. doi: 10.1093/mnras/stz491.

[33] J. Caldeira, W. L. K. Wu, B. Nord, C. Avestruz, S. Trivedi, and K. T. Story. DeepCMB: Lensing Reconstruction of the Cosmic Microwave Background with Deep Neural Networks. *Astron. Comput.*, 28:100307, 2019. doi: 10.1016/j.ascom.2019.100307.

[34] N. Krachmalnicoff and M. Tomasi. Convolutional neural networks on the healpix sphere: a pixel-based algorithm and its application to cmb data analysis. *Astronomy & Astrophysics*, 628:A129, August 2019. ISSN 1432-0746. doi: 10.1051/0004-6361/201935211. URL http://dx.doi.org/10.1051/0004-6361/201935211.

[35] J. A. Rubiño-Martín, R. Rebolo, M. Aguiar, R. Génova-Santos, F. Gómez-Reñasco, J. M. Herreros, R. J. Hoyland, C. López-Caraballo, A. E. Pelaez Santos, V. Sanchez de la Rosa, A. Vega-Moreno, T. Viera-Curbelo, E. Martínez-Gonzalez, R. B. Barreiro, F. J. Casas, J. M. Diego, R. Fernández-Cobos, D. Herranz, M. López-Caniego, D. Ortiz, P. Vielva, E. Artal, B. Aja, J. Cagigas, J. L. Cano, L. de la Fuente, A. Mediavilla, J. V. Terán, E. Villa, L. Piccirillo, R. Battye, E. Blackhurst, M. Brown, R. D. Davies, R. J. Davis, C. Dickinson, S. Harper, B. Maffei, M. McCulloch, S. Melhuish, G. Pisano, R. A. Watson, M. Hobson, K. Grainge, A. Lasenby, R. Saunders, and P. Scott. *The QUIJOTE-CMB experiment: studying the polarisation of the galactic and cosmological microwave emissions*, volume 8444 of *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, page 84442Y. 2012. doi: 10.1117/12.926581.

[36] Oludare Isaac Abiodun, Aman Jantan, Abiodun Esther Omolara, Kemi Victoria Dada, Nachaat AbdElatif Mohamed, and Humaira Arshad. State-of-the-art in artificial neural network applications: A survey. *Heliyon*, 4 (11):e00938, 2018. ISSN 2405-8440. doi: https://doi.org/10.1016/j.heliyon.2018.e00938. URL http://www.sciencedirect.com/science/article/pii/S2405844018332067.

[37] F. Rosenblatt. The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review*, pages 65–386, 1958.

[38] Chigozie Nwankpa, Winifred Ijomah, Anthony Gachagan, and Stephen Marshall. Activation functions: Comparison of trends in practice and research for deep learning, 2018.

[39] B. Widrow, Rodney Winter, and Robert Baxter. Layered neural nets for pattern recognition. *Acoustics, Speech and Signal Processing, IEEE Transactions on*, 36:1109 – 1118, 08 1988. doi: 10.1109/29.1638.

[40] Jonathan Frankle and Michael Carbin. The lottery ticket hypothesis: Training pruned neural networks. *CoRR*, abs/1803.03635, 2018. URL http://arxiv.org/abs/1803.03635.

[41] Adam Gaier and David Ha. Weight agnostic neural networks. *CoRR*, abs/1906.04358, 2019. URL http://arxiv.org/abs/1906.04358.

[42] Kurt Hornik, Maxwell Stinchcombe, Halbert White, et al. Multilayer feedforward networks are universal approximators. *Neural networks*, 2(5):359–366, 1989.

[43] Tariq Rashid. *Make Your Own Neural Network*. CreateSpace Independent Publishing Platform, North Charleston, SC, USA, 1st edition, 2016. ISBN 1530826608.

[44] Tong Zhang. Solving large scale linear prediction problems using stochastic gradient descent algorithms. In *Proceedings of the twenty-first international conference on Machine learning*, page 116, 2004.

[45] Michael A. Nielsen. Neural networks and deep learning, 2018. URL http://neuralnetworksanddeeplearning.com/.

[46] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT Press, 2016. http://www.deeplearningbook.org.

[47] David E. Rumelhart, Geoffrey E. Hinton, and Ronald J. Williams. Learning representations by back-propagating errors. *Nature*, 323(6088):533–536, 1986. ISSN 1476-4687. doi: 10.1038/323533a0. URL https://doi.org/10.1038/323533a0.

[48] U. Fuskeland, I.K. Wehus, H.K. Eriksen, and S.K. Naess. Spatial variations in the spectral index of polarized synchrotron emission in the 9-yr WMAP sky maps. *Astrophys. J.*, 790:104, 2014. doi: 10.1088/0004-637X/790/2/104.

[49] E. Carretti et al. S-band Polarization All Sky Survey (S-PASS): survey description and maps. *Mon. Not. Roy. Astron. Soc.*, 489(2):2330–2354, 2019. doi: 10.1093/mnras/stz806.

[50] A. Kogut. Synchrotron spectral curvature from 22 mhz to 23 ghz. *The Astrophysical Journal*, 753(2):110, June 2012. ISSN 1538-4357. doi: 10.1088/0004-637x/753/2/110. URL http://dx.doi.org/10.1088/0004-637X/753/2/110.

[51] Y. Akrami et al. Planck 2018 results. XI. Polarized dust foregrounds. 1 2018.

[52] David J. Schlegel, Douglas P. Finkbeiner, and Marc Davis. Maps of Dust Infrared Emission for Use in Estimation of Reddening and Cosmic Microwave Background Radiation Foregrounds. , 500(2):525–553, June 1998. doi: 10.1086/305772.

[53] Clive Dickinson, Yacine Ali-Haïmoud, Adam Barr, E. Battistelli, Aaron Bell, L. Bernstein, S. Casassus, K. Cleary, B. Draine, R. Génova-Santos, S. Harper, B. Hensley, Jaz Hill-Valler, Thiem Hoang, F. Israel, L. Jew, A. Lazarian, J. Leahy, J. Leech, and Matias Vidal. The state-of-play of anomalous microwave emission (ame) research. *New Astronomy Reviews*, 80, 02 2018. doi: 10.1016/j.newar.2018.02.001.

[54] B. T. Draine and A. Lazarian. Electric dipole radiation from spinning dust grains. *Astrophys. J.*, 508:157–179, 1998. doi: 10.1086/306387.

[55] B. T. Draine and A. Lazarian. Magnetic dipole microwave emission from dust grains. *Astrophys. J.*, 512:740, 1999. doi: 10.1086/306809.

[56] R. Génova-Santos et al. QUIJOTE scientific results – I. Measurements of the intensity and polarisation of the anomalous microwave emission in the Perseus molecular complex. *Mon. Not. Roy. Astron. Soc.*, 452(4):4169–4182, 2015. doi: 10.1093/mnras/stv1405.

[57] R. Génova-Santos et al. QUIJOTE scientific results – II. Polarisation measurements of the microwave emission in the Galactic molecular complexes

W43 and W47 and supernova remnant W44. *Mon. Not. Roy. Astron. Soc.*, 464(4):4107–4132, 2017. doi: 10.1093/mnras/stw2503.

[58] M. Remazeilles, C. Dickinson, H.K.K. Eriksen, and I.K. Wehus. Sensitivity and foreground modelling for large-scale cosmic microwave background B-mode polarization satellite missions. *Mon. Not. Roy. Astron. Soc.*, 458(2): 2032–2050, 2016. doi: 10.1093/mnras/stw441.

[59] Y. Ali-Haimoud, C. M. Hirata, and C. Dickinson. A refined model for spinning dust radiation. *Mon. Not. Roy. Astron. Soc.*, 395:1055, 2009. doi: 10.1111/j.1365-2966.2009.14599.x.

[60] Ben Thorne, Jo Dunkley, David Alonso, and Sigurd Naess. The Python Sky Model: software for simulating the Galactic microwave sky. *Mon. Not. Roy. Astron. Soc.*, 469(3):2821–2833, 2017. doi: 10.1093/mnras/stx949.

[61] Kedron Silsbee, Yacine Ali-Haïmoud, and Christopher M. Hirata. Spinning dust emission: the effect of rotation around a non-principal axis. *Monthly Notices of the Royal Astronomical Society*, 411(4):2750–2769, 03 2011. ISSN 0035-8711. doi: 10.1111/j.1365-2966.2010.17882.x. URL https://doi.org/10.1111/j.1365-2966.2010.17882.x.

[62] R. Adam et al. Planck 2015 results. X. Diffuse component separation: Foreground maps. *Astron. Astrophys.*, 594:A10, 2016. doi: 10.1051/0004-6361/201525967.

[63] E. de la Hoz, P. Vielva, R.B. Barreiro, and E. Martínez-González. On the Detection of CMB B-modes from Ground at Low Frequency. 2 2020.